Theory of solutions in the energy representation. II. Functional for the chemical potential

Matubayasi, N; Nakahara, M

JOURNAL OF CHEMICAL PHYSICS VOLUME 117, NUMBER 8 22 AUGUST 2002

Theory of solutions in the energy representation. II. Functional for the chemical potential

Nobuyuki Matubayasi[a] and Masaru Nakahara
Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

(Received 29 October 2001; accepted 31 May 2002)

An approximate functional for the chemical potential of a solute in solution is presented in the energy representation. This functional is constructed by adopting the Percus–Yevick-like approximation in the unfavorable region of the solute–solvent interaction and the hypernetted-chain-like approximation in the favorable region. The chemical potential is then expressed in terms of energy distribution functions in the solution and pure solvent systems of interest, and is given exactly to second order with respect to the solvent density and to the solute–solvent interaction. In the practical implementation, computer simulations of the solution and pure solvent systems are performed to provide the energy distribution functions constituting the approximate functional for the chemical potential. It is demonstrated that the chemical potentials of nonpolar, polar, and ionic solutes in water are evaluated accurately and efficiently from the single functional over a wide range of thermodynamic conditions. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1495850]

I. INTRODUCTION

The most fundamental quantity to describe a process in solution is the free energy change. Indeed, it governs the equilibrium and rate constants of the process. The free energy change corresponding to the insertion process of a solute in solution is the chemical potential. Once the chemical potentials are known for the species present in the initial and final states of a process of interest, the free energy change for the process can be readily evaluated. Therefore, it is of primary importance in statistical mechanics of solutions to establish a scheme to determine the chemical potential of a solute in solution.

When a set of intermolecular interaction potentials is given, the ''exact'' chemical potential of the solute under that set of potentials can be obtained, in principle, by computer simulation. The free energy perturbation and thermodynamic integration methods are commonly used techniques in free energy calculation, and evaluate the chemical potential through a gradual insertion of the solute.[1–3] In the gradual process of solute insertion, the initial and final states are the pure solvent and solution systems of interest, respectively, and the solute is partially inserted at the intermediate states.[4] The intermediate states can be chosen arbitrarily in the evaluation of the chemical potential, however, when the initial and final states are identified. Thus, an explicit reference to the intermediate states of the solute insertion process is not only computationally expensive, but is also undesirable from the conceptual viewpoint.

A molecular description of the chemical potential can be implemented by formulating a functional which expresses the chemical potential in terms only of distribution functions in the solution and pure solvent systems of interest. An approximate functional needs to be constructed in practice, however, since the exact functional involves an infinite series of many-body distribution functions.[5,6] In the method of integral equation, an approximate closure is adopted among distribution functions and often provides an analytical functional for the chemical potential.[1,7–22] Compared to the computer simulation method, the method of integral equation is efficient and can be applied to complex systems. It is disadvantageous in the account of the solution structure, on the other hand, since the integral equation for the distribution functions is not exact. Thus, the description of the chemical potential will be improved when its analytical and approximate expression is properly supplemented with computer simulation results of the distribution functions.

In the present work, we develop a functional for the chemical potential of a solute in solution. We adopt the energy representation introduced in a previous paper,[23,24] and provide an approximate but accurate functional which is expressed in terms of energy distribution functions in the solution and pure solvent systems of interest. A practical approach is then employed that computer simulations of the solution and pure solvent systems are performed to obtain the energy distribution functions constituting the functional. This approach to the chemical potential utilizes the exact solution structure and makes no reference to the intermediate states of the insertion process of the solute. Obviously, the performance of the approach is dominated by the degree of approximation involved in the functional which relates the distribution functions to the chemical potential.

The purpose of the previous paper was to formulate the theory of distribution functions in the energy representation for a dilute solution containing a single solute molecule.[4,23,24] When the energy representation is adopted, the coordinate of a solvent molecule around the solute molecule is the solute–solvent interaction energy and the solvent distribution around

[a]Author to whom correspondence should be addressed; electronic mail: nobuyuki@scl.kyoto-u.ac.jp

the solute is expressed over the one-dimensional coordinate for any type of solute–solvent interaction potential. The purpose of this work is to present a functional for the chemical potential in the energy representation and demonstrate the performance of the single functional for simple and typical model systems. We employ water as the solvent and examine the chemical potentials of nonpolar, polar, and ionic solutes over a wide range of thermodynamic conditions.

Recently, supercritical fluid receives much attention as a novel medium for chemical processes of environmental, geological, or industrial importance.[25,26] When the temperature is above the critical, a continuous variation is possible for the (solvent) density and a wide range of thermodynamic conditions is available for tuning the chemical process of interest. When a molecular fluid is to be treated at high-density conditions, the method of reference interaction site model (RISM) is useful to study the solution structure and solvation thermodynamics.[1,7–13] However, since the RISM approach with the Percus–Yevick (PY) or hypernetted-chain (HNC) closure does not give the correct low-density limit, it is not useful when the density is not high. Actually, the utility of supercritical fluid, especially of supercritical water, is often restricted to the low- to medium-density regime due to severe experimental conditions.[25–29] Therefore, to explore a wide range of thermodynamic conditions of supercritical fluid, it is necessary to employ a scheme which is accurate in a low- to medium-density fluid. The method of distribution functions in the energy representation meets this necessity. It treats each of the solute and solvent molecules as one unit, and enables a straightforward construction of a functional for the chemical potential which is exact to second order in the solvent density and in the solute–solvent interaction.[23]

The organization of the present paper is as follows. In Sec. II, a set of distribution functions are defined over the coordinate specifying the solute–solvent interaction energy and an approximate functional is presented for the chemical potential of the solute. In Sec. III, the systems to be examined are identified and the computational procedures are described. In Sec. IV, the performance of the functional for the chemical potential is assessed, and the energy distribution functions constituting the functional are characterized.

II. THEORY

The purpose of this section is to construct an approximate functional for the chemical potential in terms of energy distribution functions. The construction consists of two subsections. In Sec. II A, we present a brief introduction of the energy representation developed in Ref. 23, and define the energy distribution functions necessary for formulating the functional for the chemical potential. In Sec. II B, we describe the approximation and give the explicit expression for the functional.

A. Distribution functions in the energy representation

The system of our interest is a dilute solution containing a single solute molecule.[4] The solute molecule is fixed at the (arbitrarily chosen) origin with an (arbitrarily chosen) fixed orientation. In the present work, it is supposed that the interaction between the solute molecule and a solvent molecule is pairwise and depends only on the configuration of the solvent molecule relative to the solute molecule. The notations in this paper are the same as those adopted in Ref. 23. The complete set of the position and orientation of a solvent molecule is called the full coordinate and is denoted collectively by $\mathbf{x}$.[30] The solute–solvent interaction potential of interest is v and is fixed at the outset in our developments. The energy representation is then introduced by adopting the value of v as the coordinate ϵ of the solvent molecule. The superscripts f and e are attached, respectively, to emphasize that a function is represented over the full coordinate $\mathbf{x}$ and over the energy coordinate ϵ. Obviously, the solute–solvent interaction v of interest satisfies $v^e(\epsilon)=\epsilon$ by definition.

The instantaneous distribution $\hat{\rho}^e$ is defined in the energy representation as

$$\hat{\rho}^e(\epsilon)=\sum_i \delta(v^f(\mathbf{x}_i)-\epsilon), \tag{1}$$

where $\mathbf{x}_i$ is the full coordinate of the ith solvent molecule. In the energy representation, the specification of v is necessary to introduce $\hat{\rho}^e$. The average distribution ρ^e of the v value is correspondingly expressed in the presence of a solute–solvent interaction u as

$$\rho^e(\epsilon;u)=\langle\hat{\rho}^e(\epsilon)\rangle_u, \tag{2}$$

where $\langle\cdots\rangle_u$ denotes the ensemble average taken in the solution with the solute–solvent interaction u. Note that v serves to construct the solvent coordinate ϵ and does not identify the system in which the average is taken (unless $u=v$). The solute–solvent interaction potential u, on the other hand, specifies the solution and determines the system in which the averaging is carried out. The correlation matrix χ^e in the energy representation is further defined as

$$\chi^e(\epsilon,\eta;u)=\langle\hat{\rho}^e(\epsilon)\hat{\rho}^e(\eta)\rangle_u-\langle\hat{\rho}^e(\epsilon)\rangle_u\langle\hat{\rho}^e(\eta)\rangle_u. \tag{3}$$

χ^e is positive definite and invertible, and is symmetric with respect to the two arguments of the solvent coordinate.

The energy coordinate ϵ is introduced with respect to the solute–solvent interaction v of interest. The correspondence from the solute–solvent interaction potential to the solvent distribution around the solute is then one-to-one in the energy representation, as shown in Ref. 23, by focusing on a set of potentials u which are constant over equienergy surfaces of v. A potential function u contained in this set may be considered to be defined over the energy coordinate ϵ and can be expressed as $u^e(\epsilon)$. In the following developments, we restrict our attention to the solute–solvent interaction potentials u in the form of $u^e(\epsilon)$.

To describe the approximation in Sec. II B, it is useful to employ the indirect part of the potential of mean force between the solute and solvent molecules. The indirect part w^e of the solute–solvent potential of mean force in the energy representation is defined in the presence of a solute–solvent interaction u as

$$w^e(\epsilon;u)=-k_BT\log\left(\frac{\rho^e(\epsilon;u)}{\rho^e(\epsilon;0)}\right)-u^e(\epsilon), \tag{4}$$

where k_B is the Boltzmann constant, T is the temperature, and $\rho^e(\epsilon;0)$ is the distribution function in the pure solvent

system ($u=0$). w^e reflects the many-body effects in the solute–solvent correlation and vanishes in the limit of zero solvent density. It should be noted that the definitions given by Eqs. (1), (2), (3), and (4) are identical to Eqs. (3), (4), (15), and (7) of Ref. 23, respectively.

In Sec. II B, we express the chemical potential in terms of the distribution functions in the solution of interest ($u=v$) and in the pure solvent ($u=0$). In this connection, it is convenient to introduce simplified notations through

$$\begin{aligned}
&\rho^e(\epsilon)=\rho^e(\epsilon;v)\\
&\rho_0^e(\epsilon)=\rho_0^e(\epsilon;0)\\
&\chi_0^e(\epsilon,\eta)=\chi^e(\epsilon,\eta;0)\\
&w^e(\epsilon)=w^e(\epsilon;v)\\
&w_0^e(\epsilon)=-k_BT\int d\eta\left(\frac{\delta(\epsilon-\eta)}{\rho_0^e(\epsilon)}\right.\\
&\qquad\left.-(\chi_0^e)^{-1}(\epsilon,\eta)\right)(\rho^e(\eta)-\rho_0^e(\eta)).
\end{aligned}\tag{5}$$

In Eq. (5), the notational structure of w_0^e is different from the others. Actually, $w^e(\epsilon;u)$ vanishes in the pure solvent ($u=0$). We introduced w_0^e in Eq. (5) because the PY and HNC approximations in the energy representation are given, respectively, by

$$w^e(\epsilon)=-k_BT\log(1-\beta w_0^e(\epsilon))\tag{6}$$

and

$$w^e(\epsilon)=w_0^e(\epsilon),\tag{7}$$

where β is the inverse of k_BT.[23] Furthermore, when the pure solvent is homogeneous and isotropic,[31] $\rho_0^e(\epsilon)$ reduces to

$$\rho_0^e(\epsilon)=\rho^n\int d\mathbf{x}\delta(v^f(\mathbf{x})-\epsilon),\tag{8}$$

where ρ^n is the number density of the pure solvent and the integral over $\mathbf{x}$ is the density of states for the solute–solvent interaction v of interest.

B. Construction of the chemical potential

In order to treat the thermodynamics of solvation on a definite basis, it is necessary to specify the process of solute insertion. We consider the insertion process of the solute at the (arbitrarily chosen) fixed origin with an (arbitrarily chosen) fixed orientation. The free energy change for this process is the excess chemical potential $\Delta\mu$ of the solute.[1] $\Delta\mu$ involves only the contribution from the potential energy, and the ideal (kinetic) contribution is excluded at the outset. In the insertion process, the solute–solvent interaction is gradually turned on according to the coupling parameter λ ($0\leqslant\lambda\leqslant1$). When $\lambda=0$, there is no interaction between the solute and solvent and the system is the pure solvent.[4] When $\lambda=1$, the solute interacts with the solvent at full coupling under the solute–solvent interaction potential v of interest. The gradual insertion of the solute is described in the energy representation by a family of solute–solvent interaction potentials u_λ defined over the coupling parameter λ and the energy coordinate ϵ. Of course, it is imposed that

$$\begin{aligned}
&u_0^e(\epsilon)=0\\
&u_1^e(\epsilon)=v^e(\epsilon)=\epsilon.
\end{aligned}\tag{9}$$

The excess chemical potential $\Delta\mu$ of the solute of interest is expressed as

$$\Delta\mu=\int_0^1 d\lambda\int d\epsilon\frac{\partial u_\lambda^e(\epsilon)}{\partial\lambda}\rho^e(\epsilon;u_\lambda).\tag{10}$$

Equation (10) is the charging formula in the energy representation and is exact for any choice of u_λ.[23] In the present treatment, u_λ is chosen so that $\rho^e(\epsilon;u_\lambda)$ varies linearly against λ and that

$$\rho^e(\epsilon;u_\lambda)=\lambda\rho^e(\epsilon)+(1-\lambda)\rho_0^e(\epsilon)\tag{11}$$

holds at each ϵ. The unique existence of the u_λ satisfying Eq. (11) is assured by the one-to-one correspondence between the set of solute–solvent interaction potentials defined over the energy coordinate and the corresponding set of solvent distribution functions around the solute in the energy representation.[23] Under the particular choice of u_λ identified by Eq. (11), the excess chemical potential $\Delta\mu$ is written as

$$\begin{aligned}
\Delta\mu=-k_BT\int d\epsilon\Big[&(\rho^e(\epsilon)-\rho_0^e(\epsilon))\\
&+\beta w^e(\epsilon)\rho^e(\epsilon)-\left(\beta\int_0^1 d\lambda w^e(\epsilon;u_\lambda)\right)\\
&\times(\rho^e(\epsilon)-\rho_0^e(\epsilon))\Big].
\end{aligned}\tag{12}$$

Furthermore,

$$w_0^e(\epsilon)=\left.\frac{\partial w^e(\epsilon;u_\lambda)}{\partial\lambda}\right|_{\lambda=0}\tag{13}$$

is valid by virtue of Eq. (5). It should be noted that Eqs. (12) and (13) are exact when $\rho^e(\epsilon;u_\lambda)$ undergoes a linear variation against the coupling parameter λ.

When the PY approximation in the energy representation is adopted along the λ variation according to Eq. (11), the indirect part $w^e(\epsilon;u_\lambda)$ of the solute–solvent potential of mean force is expressed as

$$\begin{aligned}
-\beta w^e(\epsilon;u_\lambda)&=\log(1+\lambda(\exp(-\beta w^e(\epsilon))-1))\\
&=\log(1-\lambda\beta w_0^e(\epsilon)).
\end{aligned}\tag{14}$$

In this case, the λ integral of $w^e(\epsilon;u_\lambda)$ in Eq. (12) reduces to

$$\begin{aligned}
\beta\int_0^1 d\lambda w^e(\epsilon;u_\lambda)&=\beta w^e(\epsilon)+1+\frac{\beta w^e(\epsilon)}{\exp(-\beta w^e(\epsilon))-1}\\
&=-\log(1-\beta w_0^e(\epsilon))+1\\
&\quad+\frac{\log(1-\beta w_0^e(\epsilon))}{\beta w_0^e(\epsilon)}.
\end{aligned}\tag{15}$$

In the HNC approximation, $w^e(\epsilon;u_\lambda)$ is simply given by

$$w^e(\epsilon;u_\lambda)=\lambda w^e(\epsilon)=\lambda w_0^e(\epsilon), \tag{16}$$

and its integral is

$$\int_0^1 d\lambda\, w^e(\epsilon;u_\lambda)=\tfrac{1}{2}w^e(\epsilon) = \tfrac{1}{2}w_0^e(\epsilon). \tag{17}$$

Of course, the PY and HNC approximations are not exact and Eq. (6) or (7) cannot be assumed in general. In the present work, we adopt the exact values of $w^e(\epsilon)$ and $w_0^e(\epsilon)$ and construct an approximate expression for the λ integral of $w^e(\epsilon;u_\lambda)$ in Eq. (12).

Our assumption for the λ integral is expressed as

$$\beta\int_0^1 d\lambda\, w^e(\epsilon;u_\lambda)=\alpha(\epsilon)F(\epsilon) +(1-\alpha(\epsilon))F_0(\epsilon), \tag{18}$$

where $F(\epsilon)$ and $F_0(\epsilon)$ are functions only of $w^e(\epsilon)$ and $w_0^e(\epsilon)$, respectively, and $\alpha(\epsilon)$ is a weight factor which is unity when $\rho^e(\epsilon)\gg\rho_0^e(\epsilon)$ and vanishes when $\rho^e(\epsilon)\ll\rho_0^e(\epsilon)$. In this assumption, the λ integral reflects the properties of the solution ($\lambda=1$) in the favorable region of solvation, where $\rho^e(\epsilon)\gg\rho_0^e(\epsilon)$ holds. On the other hand, the λ integral reflects only the properties of the pure solvent ($\lambda=0$) in the unfavorable region, where $\rho^e(\epsilon)\ll\rho_0^e(\epsilon)$ is valid. We determine the explicit forms of $F(\epsilon)$ and $F_0(\epsilon)$ with the help of Eqs. (15) and (17). We adopt the PY form in the unfavorable region, where $w^e(\epsilon)$ and $w_0^e(\epsilon)$ are mostly negative, and the HNC form in the favorable region, where $w^e(\epsilon)$ and $w_0^e(\epsilon)$ are mostly positive.[32] Our forms of $F(\epsilon)$ and $F_0(\epsilon)$ are then given by

$$F(\epsilon)=\begin{cases}\beta w^e(\epsilon)+1+\dfrac{\beta w^e(\epsilon)}{\exp(-\beta w^e(\epsilon))-1} & (\text{when } w^e(\epsilon)\leqslant 0)\\ \tfrac{1}{2}\beta w^e(\epsilon) & (\text{when } w^e(\epsilon)\geqslant 0)\end{cases} \tag{19}$$

$$F_0(\epsilon)=\begin{cases}-\log(1-\beta w_0^e(\epsilon))+1+\dfrac{\log(1-\beta w_0^e(\epsilon))}{\beta w_0^e(\epsilon)} & (\text{when } w_0^e(\epsilon)\leqslant 0)\\ \tfrac{1}{2}\beta w_0^e(\epsilon) & (\text{when } w_0^e(\epsilon)\geqslant 0)\end{cases}. \tag{20}$$

It should be noted that when Eq. (6) holds for a negative $w^e(\epsilon)$ and Eq. (7) holds for a positive $w^e(\epsilon)$, $F(\epsilon)$ in Eq. (19) is equal to $F_0(\epsilon)$ in Eq. (20). Furthermore, the approximate value of the excess chemical potential $\Delta\mu$ determined from Eqs. (12), (18), (19), and (20) is exact to second order in the solvent density and in the solute–solvent interaction potential for any choice of the weight factor $\alpha(\epsilon)$ in Eq. (18). In this work, we set $\alpha(\epsilon)$ to the form of

$$\alpha(\epsilon)=\begin{cases}1 & (\text{when } \rho^e(\epsilon)\geqslant\rho_0^e(\epsilon))\\ 1-\left(\dfrac{\rho^e(\epsilon)-\rho_0^e(\epsilon)}{\rho^e(\epsilon)+\rho_0^e(\epsilon)}\right)^2 & (\text{when } \rho^e(\epsilon)\leqslant\rho_0^e(\epsilon))\end{cases}. \tag{21}$$

Equations (12), (18), (19), (20), and (21) constitute the approximate functional for the excess chemical potential in terms of the energy distribution functions in the solution and pure solvent systems.[31]

The assumed form given by Eq. (18) is of practical use from the computational viewpoint. In the favorable region of solvation, $\rho^e(\epsilon)$ is not small and $w^e(\epsilon)$ can be obtained with good precision. In the unfavorable region, on the other hand, it is difficult to determine $w^e(\epsilon)$ according to Eq. (4) since $\rho^e(\epsilon)$ is small and ϵ is large. Actually, $\rho_0^e(\epsilon)$ is often substantial in the unfavorable region, where the solute molecule overlaps significantly with solvent molecules. Thus, when ϵ is large, $\chi_0^e(\epsilon,\eta)$ can be easily calculated in a computer simulation of the pure solvent system and Eq. (5) provides $w_0^e(\epsilon)$ with good precision.[33]

It is often the case that the solute–solvent interaction potential v of interest is essentially of finite range and that its long-range part may be safely neglected to account for the physics of the solution. When v is zero outside a finite region Ω, it is useful to employ a reduced form $\hat{\rho}_\Omega^e$ of instantaneous distribution defined as

$$\hat{\rho}_\Omega^e(\epsilon)=\sum_{i\in\Omega}\delta(v^f(\mathbf{x}_i)-\epsilon), \tag{22}$$

where the sum is taken over the solvent molecules contained within Ω. An approximate procedure for the excess chemical potential can then be reformulated straightforwardly by replacing $\hat{\rho}^e$ with $\hat{\rho}_\Omega^e$ in the developments of the present section. The potential functions treated in Secs. III and IV are actually of finite range. We thus employ the version for a solute–solvent interaction of finite range in the practical implementations, although we refer to the expressions without the subscript Ω for notational simplicity.

III. PROCEDURES

This section consists of two subsections. In Sec. III A we specify the potential functions for the solute and solvent mol-

ecules. In Sec. III B we describe the simulation procedures for the solution and pure solvent systems of interest.

A. System

In this work, the solvent is water. The solution is dilute and contains a single solute molecule. The thermodynamic state of each system of interest is specified by the water density and temperature. Four thermodynamic states are examined. One is an ambient state of 1.0 g/cm^3 and 25 °C and the others are supercritical states of 1.0, 0.6, and 0.2 g/cm^3 and 400 °C. The water molecule is treated as rigid and nonpolarizable and the SPC/E model is adopted as the intermolecular potential function between water molecules.[34,35] The intermolecular interaction between a pair of water molecules is spherically truncated at 9.0 Å on the basis of the oxygen–oxygen distance.[36]

The solute species treated in the present work are nonpolar, polar, and ionic. Each solute species interacts with the water molecule through

$$v(\mathbf{x})=S(\mathbf{x})\sum_{i,j}\left(4\epsilon_{ij}\left\{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right\}+\frac{q_iq_j}{r_{ij}}\right), \tag{23}$$

where $v(\mathbf{x})$ is the potential function of interest determined by the configuration $\mathbf{x}$ of the water molecule relative to the solute molecule and $S(\mathbf{x})$ controls the truncation of the solute–water interaction.[36] In the sum of Eq. (23), i and j refer to solute and solvent interaction sites, respectively. The first term in the sum expresses the Lennard-Jones interaction at the distance r_{ij} between the solute and solvent sites, and ϵ_{ij} and σ_{ij} are the energy and length parameters, respectively. The second term in the sum corresponds to the Coulombic interaction, and q_i and q_j are the charges on the solute and solvent sites, respectively.

The nonpolar solutes treated are methane, ethane, and a purely repulsive derivative of methane. The methane molecule is treated as a united atom and interacts with a water molecule through the Lennard-Jones interaction depending only on its distance from the oxygen site. The Coulombic interaction in Eq. (23) is absent. The Lennard-Jones parameters are $\sigma_{\mathrm{Me-O}}=3.4475$ Å and $\epsilon_{\mathrm{Me-O}}=0.2134$ kcal/mol.[37,38] The spherical truncation is applied at 9.0 Å, where the potential is shifted according to the standard prescription.[2] A fused pair of methane molecules whose distance is fixed at 1.5 Å is referred to as ethane. The potential between ethane and water is truncated when the closer of the two methane–oxygen pairs is separated by more than 9.0 Å. In addition to methane and ethane, we examine a purely repulsive solute derived from the Weeks–Andersen–Chandler (WCA) separation of the methane–water potential.[39] This solute is called ''WCA–methane'' in the following, and its interaction with water is spherically truncated at the methane–oxygen distance corresponding to the minimum of the methane–water potential.

Water, methanol, and ethanol are employed as polar solutes. The interaction between the ''solute'' water molecule and the ''solvent'' water molecule is the same as that described above for the solvent water. Methanol and ethanol are treated as three- and four-site molecules, respectively, by adopting the united-atom approximation for the methyl and methylene groups. The conformation of ethanol is fixed at the trans. In the OH moiety, the Lennard-Jones interaction is absent at the hydrogen site and acts only between the oxygen sites of the alcohol and water.[35] The Lennard-Jones parameters and partial charges for methanol and ethanol are taken from the OPLS parameter set,[40] and the standard Lorentz–Berthelot combining rule is used for constructing the Lennard-Jones parts of the potential functions between the alcohols and water.[1] The interactions between the alcohols and water are truncated spherically at 9.0 Å with respect to the distance between the oxygen sites.

The ionic solutes examined are sodium and chloride ions. In this case, the interaction parameter set by Pettitt and Rossky is adopted[35,41] and $S(\mathbf{x})$ in Eq. (23) is given by

$$S(\mathbf{x})=\begin{cases}1 & (r<r_c-\Delta)\\ \dfrac{(r-r_c)^3-3\Delta^2(r-r_c)+2\Delta^3}{4\Delta^3} & (r_c-\Delta\leqslant r\leqslant r_c+\Delta),\\ 0 & (r_c+\Delta<r)\end{cases} \tag{24}$$

where r is the distance between the ion and the oxygen site of water.[42] The parameters r_c and Δ are set to 9.0 and 0.5 Å, respectively. In addition, NaCl is employed as an ion pair. It is an associated pair of sodium and chloride ions, and the interactions of the constituent ions with water are identical to those described above for the individual ions. The distance between the sodium and chloride ions in the pair is fixed at 2.8 Å, which corresponds to the Na–Cl distance in the crystal structure.

B. Simulation

The distribution functions $\rho^e(\epsilon)$, $\rho_0^e(\epsilon)$, and $\chi_0^e(\epsilon,\eta)$ given in Eq. (5) are the inputs needed to evaluate the excess chemical potential $\Delta\mu$ of the solute through Eqs. (12), (18), (19), (20), and (21). In order to obtain $\rho^e(\epsilon)$, a Monte Carlo simulation was conducted for the solution system. In each Monte Carlo simulation, one solute molecule of interest and 647 water molecules were located in a cubic unit cell and the

standard Metropolis sampling scheme was implemented in the canonical ensemble.[2] The Monte Carlo simulation was performed for 10 K passes, where one pass corresponds to the generation of 648 configurations. The periodic boundary condition was employed in the minimum image convention, and the method of preferential sampling was not used. The instantaneous distribution $\hat{\rho}^e(\epsilon)$ defined by Eq. (1) [actually, its reduced form defined by Eq. (22)] was sampled every pass and was averaged through Eq. (2) over 10 K configurations of the solution system. In order to obtain $\rho_0^e(\epsilon)$ and $\chi_0^e(\epsilon,\eta)$, on the other hand, a Monte Carlo simulation was carried out for the pure solvent system. In this case, the standard Metropolis sampling scheme in the canonical ensemble was implemented by locating 648 water molecules in a cubic unit cell. The size of the unit cell was identical to that of the corresponding simulation of the solution system consisting of one solute molecule and 647 water molecules. The simulation length was 5 K passes, and the boundary condition was the same as that for the solution system. When an (instantaneous) configuration of the pure solvent system is sampled to construct the instantaneous distribution $\hat{\rho}^e(\epsilon)$, the solute molecule of interest is inserted as a test particle at a random position in the unit cell with a random orientation. $\hat{\rho}^e(\epsilon)$ is then the histogram of the interaction potential energies between the solute molecule inserted and the solvent molecules, and is averaged to give $\rho_0^e(\epsilon)$ and $\chi_0^e(\epsilon,\eta)$ according to Eqs. (3) and (8). The configuration of the pure solvent system was sampled every 100 passes, and the insertion of the solute was performed 2 K times at each solvent configuration sampled. $\rho_0^e(\epsilon)$ and $\chi_0^e(\epsilon,\eta)$ were thus obtained from the averaging of $\hat{\rho}^e(\epsilon)$ over 100 K insertions in a single simulation for 5 K passes. It should be noted that the implementation of the procedure to obtain $\rho_0^e(\epsilon)$ and $\chi_0^e(\epsilon,\eta)$ does not affect the sequence of configurations generated in the simulation.[43]

One set of simulations to approximately evaluate the excess chemical potential $\Delta\mu$ consists of two simulations. One is of the solution system and the other is of the pure solvent system.[44] As described above, the length of one simulation is 10 K and 5 K passes for the solution and pure solvent systems, respectively. For each solute and at each thermodynamic state, we actually performed six sets of simulations and obtained six values of the excess chemical potential denoted by $\Delta\mu_i$ $(i=1,\cdots,6)$. We then estimated the average by

$$\frac{1}{n}\sum_i \Delta\mu_i \tag{25}$$

and the error at 95% confidence level through

$$\frac{2}{\sqrt{n}}\sqrt{\frac{1}{n-1}\left\{\sum_i \Delta\mu_i^2-\frac{1}{n}\left(\sum_i \Delta\mu_i\right)^2\right\}}, \tag{26}$$

where $n=6$.

Each solute–solvent interaction potential treated in the present work is continuous over its range. In other words, the corresponding energy coordinate is continuous. In this case, a discretization of the energy coordinate is needed for the numerical realization of the energy distribution functions. The scheme of discretization is described in detail in the Appendix.

In order to assess the accuracy of an approximate procedure for a thermodynamic quantity under a given set of potential functions, the exact evaluation is required for the thermodynamic quantity under the same set of potential functions.[1,2] The exact values of the excess chemical potentials of the nonpolar solutes were calculated by means of the particle insertion method. The calculations were performed for methane and ethane in Refs. 45 and 46,[47] and for WCA–methane in the computer simulations of pure solvent described above in the present paper. For the other solutes, the free energy perturbation method was implemented in this work. In each free energy calculation, a Monte Carlo simulation was carried out using the standard Metropolis sampling scheme in the canonical ensemble. One solute molecule and 647 water molecules were then located in a cubic unit cell, and the size of the unit cell was taken to be identical to that of the corresponding simulations of the solution and pure solvent systems. The periodic boundary condition in the minimum image convention was employed, and the preferential sampling method was not used.

In the free energy perturbation method, the solute–solvent interaction is controlled by the coupling parameter λ $(0\leqslant\lambda\leqslant1)$. When $0\leqslant\lambda\leqslant1/3$, the Lennard-Jones term in Eq. (23) was turned on, as motivated by Zacharias *et al.*,[48] according to

$$12\lambda\epsilon_{ij}\left\{\frac{{\sigma_{ij}}^{12}}{(3\lambda r_{ij}^2+(1-3\lambda)\sigma_{ij}^2)^6}-\frac{\sigma_{ij}^6}{(3\lambda r_{ij}^2+(1-3\lambda)\sigma_{ij}^2)^3}\right\}. \tag{27}$$

In this region of λ, the (partial) charges on the solute molecule were set to zero and the Coulombic term was absent. When $1/3\leqslant\lambda\leqslant1$, on the other hand, the Lennard-Jones term was at full coupling in the form of Eq. (23) and the Coulombic term was linearly scaled through

$$\frac{3}{2}\left(\lambda-\frac{1}{3}\right)\frac{q_iq_j}{r_{ij}}. \tag{28}$$

In our calculation, the coupling parameter λ was varied in 90 steps with the equally spaced intervals. At each value of λ, the system was equilibrated for 5 K passes and the free energy change to the system at the next λ was calculated for 5 K passes. The variation of λ from 0 to 1 corresponds to the creation of the solute molecule and the reverse variation from 1 to 0 corresponds to the annihilation. We performed three sets of free energy perturbation calculations for both the creation and annihilation processes. Six values were then obtained for the excess chemical potential, and the average and error were estimated by Eqs. (25) and (26), respectively.

IV. RESULTS AND DISCUSSION

The approximate values of the excess chemical potentials are listed in Table I and are compared to the exact values obtained from the free energy calculations. The overall

TABLE I. Excess chemical potential in the unit of kcal/mol.[a]

Solute	Thermodynamic state			
	1.0 g/cm^3 and 25 °C	1.0 g/cm^3 and 400 °C	0.6 g/cm^3 and 400 °C	0.2 g/cm^3 and 400 °C
WCA–methane	6.8±0.1 (6.5±0.4)	13.0±0.1 (15.2±0.8)	5.0 (4.7±0.1)	1.2 (1.1)
methane[b]	3.0±0.2 (2.8)	9.4±0.1 (10.7)	3.1 (2.8)	0.6 (0.6)
ethane[b]	2.6±0.3 (1.9)	11.3±0.1 (13.2)	3.2±0.1 (2.8)	0.6 (0.5)
water	−8.2±0.7 (−6.9±0.7)	−1.5±0.4 (−0.7±0.3)	−3.3±0.2 (−3.1±0.1)	−2.4±0.2 (−2.3±0.2)
methanol	−4.4±0.5 (−3.9±1.3)	4.5±0.4 (6.0±0.3)	−0.6±0.3 (−0.8±0.1)	1.2±0.1 (−1.3±0.1)
ethanol	−4.4±0.4 (−3.0±1.2)	8.0±0.3 (10.9±0.3)	0.4±0.1 (0.0±0.2)	−0.9±0.1 (−1.0±0.1)
Na^+	−96.0±1.9 (−100.5±0.7)	−85.9±0.9 (−90.2±0.1)	−80.2±0.7 (−83.1±0.3)	−67.5±0.9 (−74.5±0.6)
Cl^-	−69.3±1.6 (−64.2±1.4)	−54.0±0.9 (−51.5±0.3)	−53.9±0.6 (−54.4±0.2)	−45.5±0.9 (−46.7±0.4)
NaCl	−82.9±1.8 (−78.6±1.3)	−61.4±0.9 (−59.7±0.4)	−58.7±0.9 (−59.1±0.4)	−47.5±1.2 (−50.5±0.4)

[a]Each entry contains the approximate and exact values and the exact value is in parenthesis. Each value is rounded to a multiple of 0.1 kcal/mol.
[b]The exact values for methane and ethane are taken from Refs. 45 and 46. Their errors are smaller than 0.1 kcal/mol since the simulation lengths in Refs. 45 and 46 were long (on the order of 10^6 to 10^7 passes) and the particle insertion method was applicable in the free energy calculations.

agreement between the approximate and exact values is good. The agreement is particularly notable at the low- and medium-density states of 0.6 and 0.2 g/cm^3 and 400 °C, while RISM is not competent for a low- to medium-density molecular fluid. At the high-density states of 1.0 g/cm^3 and 25 and 400 °C, the approximate values for the neutral, polar solutes are systematically seen to be more favorable (smaller) than the corresponding exact values. When the solute is ionic, the density at the state of 0.2 g/cm^3 and 400 °C is not yet ''low enough'' in the sense, for example, that the hydration number at that state is comparable to the numbers at ambient states.[49,50] Even in this case, our approximate procedure is effective in determining the excess chemical potential. The exact excess chemical potentials of water at 1.0 g/cm^3 and 400 °C and of methanol and ethanol at 0.6 g/cm^3 and 400 °C are rather small in magnitude. These behaviors are caused by the balance between the favorable and unfavorable contributions of the solute–solvent interactions to the excess chemical potentials, and are well reproduced by the approximate method in the present paper. Therefore, the single functional given by Eqs. (12), (18), (19), (20), and (21) provides an accurate and efficient route to the excess chemical potential for various types of solutes over a wide range of thermodynamic conditions.

In the approximate evaluation of the excess chemical potential through Eqs. (12), (18), (19), (20), and (21), the inputs from the computer simulation are the distribution functions $\rho^e(\epsilon)$, $\rho_0^e(\epsilon)$, and $\chi_0^e(\epsilon,\eta)$ given in Eq. (5). It is then insightful to note the connection of these energy distribution functions to the corresponding distribution functions in the full coordinate representation and illustrate the intermolecular correlation over the energy coordinate. In the intermolecular interaction potentials treated in the present work, $\rho_0^e(\epsilon)$ is actually trivial with respect to the intermolecular correlation. Indeed, when ρ^n is the number density of the solvent, Eq. (8) shows that $\rho_0^e(\epsilon)/\rho^n$ is simply the density of states for the solute–solvent interaction and is independent of the thermodynamic state. On the other hand, $\rho^e(\epsilon)$ provides an energy-represented description of the solute–solvent correlation in the solution system. In the full coordinate representation, the solute–solvent correlation is described by a correlation function $g_{uv}^f(\mathbf{x})$ which is defined so that $\rho^n g_{uv}^f(\mathbf{x})$ is the (number) density of the solvent at the full coordinate $\mathbf{x}$ relative to the solute molecule fixed at the origin with a fixed orientation. It is then easy to see from the definition given in Sec. II A of this paper and Ref. 23 that

$$\frac{\rho^e(\epsilon)}{\rho^n}=\int d\mathbf{x}\,\delta(v^f(\mathbf{x})-\epsilon)g_{uv}^f(\mathbf{x}), \tag{29}$$

where $v^f(\mathbf{x})$ is the solute–solvent interaction of interest in the full coordinate representation. According to Eq. (29), $\rho^e(\epsilon)/\rho^n$ corresponds to an average of g_{uv}^f over an equienergy surface. When the system is the pure solvent and is homogeneous and isotropic, the solvent–solvent correlation function $g_{vv,0}^f(\mathbf{x},\mathbf{y})$ is introduced in the full coordinate representation by taking $(\rho^n)^2 g_{vv,0}^f(\mathbf{x},\mathbf{y})$ to be the two-body density of the pure solvent system.[1,51] This function is then related to $\chi_0^e(\epsilon,\eta)$ in the energy representation through

$$\begin{aligned}\chi_0^e(\epsilon,\eta)=&\,\delta(\epsilon-\eta)\rho_0^e(\epsilon)\\&+(\rho^n)^2\int d\mathbf{x}\int d\mathbf{y}\,\delta(v^f(\mathbf{x})-\epsilon)\,\delta(v^f(\mathbf{y})-\eta)\\&\times(g_{vv,0}^f(\mathbf{x},\mathbf{y})-1)\end{aligned} \tag{30}$$

by the definition presented in Sec. II A of this paper and Ref. 23. Thus, $\chi_0^e(\epsilon,\eta)$ corresponds to an average of $g_{vv,0}^f$ and provides a reduced description of the two-body correlation in the pure solvent system.

In Figs. 1(a), 1(b), and 1(c), we show $\rho^e(\epsilon)/\rho^n$ and $\rho_0^e(\epsilon)/\rho^n$ for methane, water, and sodium ion, respectively, at the four thermodynamic states examined in the present work. For each solute, $\rho^e(\epsilon)$ vanishes when the solute–solvent interaction is strongly unfavorable and the argument ϵ is large. On the other hand, $\rho_0^e(\epsilon)$ cannot be set to zero at large ϵ due to the overlap of the solvent molecules with the randomly inserted solute molecule. Actually, $\rho_0^e(\epsilon)$ is integrated over the large ϵ region in Eq. (12) to account for the excluded volume effect on the excess chemical potential $\Delta\mu$. When ϵ approaches zero, both $\rho^e(\epsilon)$ and $\rho_0^e(\epsilon)$ become large. This is simply a reflection of the fact that the solvent molecule interacting weakly with the solute molecule ($\epsilon\approx 0$) is large in

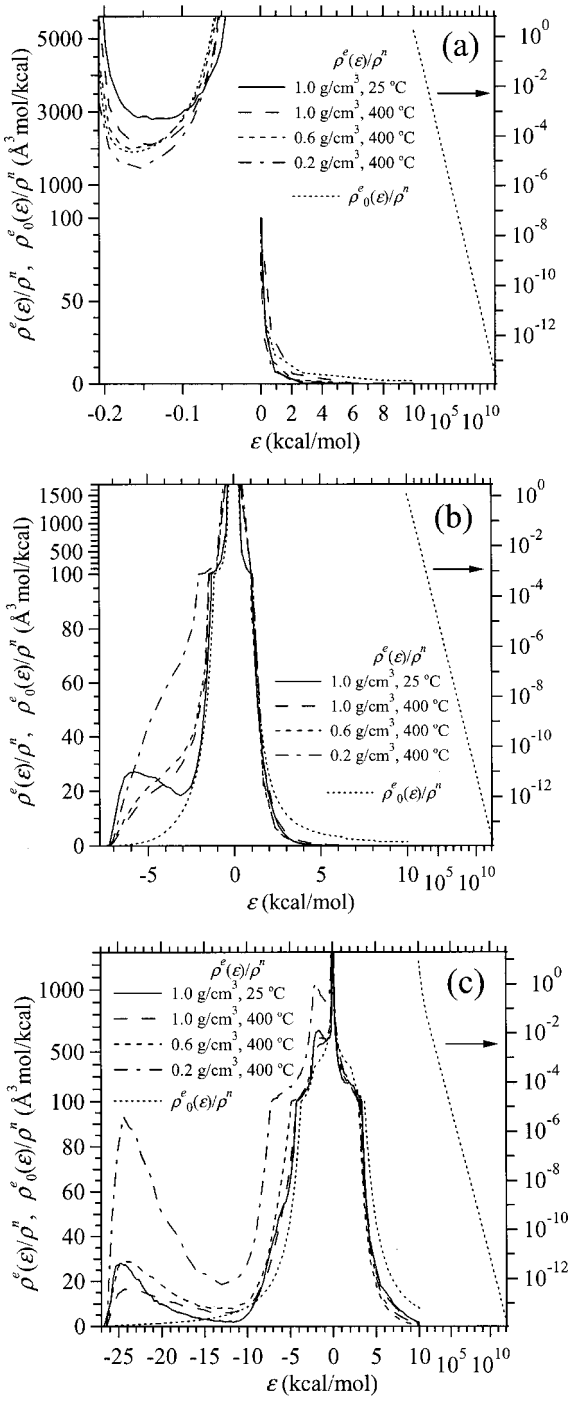

FIG. 1. The energy distribution functions $\rho^e(\epsilon)/\rho^n$ and $\rho^e_0(\epsilon)/\rho^n$ for: (a) methane, (b) water, and (c) sodium ion at an ambient state of 1.0 g/cm^3 and 25 °C and at supercritical states of 1.0, 0.6, and 0.2 g/cm^3 and 400 °C, where ρ^n is the number density of the solvent. $\rho^e_0(\epsilon)/\rho^n$ is determined only by the solute–solvent interaction potential and is independent of the thermodynamic state. The abscissa is graduated linearly when $\epsilon \lesssim 10$ kcal/mol and logarithmically when $\epsilon \gtrsim 10$ kcal/mol. In (a), the graduation is also changed at $\epsilon = 0$ kcal/mol. When $\epsilon \lesssim 10$ kcal/mol, the ordinate refers to the left. It is linearly graduated, whereas the scale is changed at 100 Å^3 mol/kcal. When $\epsilon \gtrsim 10$ kcal/mol, the ordinate refers to the right and only $\rho^e_0(\epsilon)/\rho^n$ is shown.

number. With respect to the sign of ϵ, however, the increasing tendency toward $\epsilon = 0$ depends on the type of solute. When the solute is nonpolar, the solute–solvent interaction in the present work is always nonpositive at large distances. It is then seen in Fig. 1(a) that $\rho^e(\epsilon)$ and $\rho^e_0(\epsilon)$ are large at small $|\epsilon|$ only when ϵ is negative. When the solute is polar or ionic, in contrast, both positive and negative values are possible for the solute–solvent interaction at large distances. Indeed, Figs. 1(b) and 1(c) show that $\rho^e(\epsilon)$ and $\rho^e_0(\epsilon)$ at small $|\epsilon|$ are large for both positive and negative ϵ.

When the solute is methane, the solute–solvent interaction described in Sec. III A is the Lennard-Jones interaction depending only on the distance between the methane and oxygen site of water, and involves a minimum value of $-\epsilon_{\mathrm{Me-O}}$. In this case, $\rho^e(\epsilon)$ and $\rho^e_0(\epsilon)$ diverge at $\epsilon = -\epsilon_{\mathrm{Me-O}}$. From Eq. (29), however, the divergence is proportional to $1/\sqrt{\epsilon+\epsilon_{\mathrm{Me-O}}}$ and no singularity is present in the ϵ integral of Eq. (12). When the temperature is elevated from 25 to 400 °C at a constant density of 1.0 g/cm^3, $\rho^e(\epsilon)$ is observed in Fig. 1(a) to decrease at $\epsilon \lesssim -0.1$ kcal/mol and increase at $\epsilon \gtrsim -0.1$ kcal/mol. This observation actually corresponds to the complementary observation for the radial distribution function. In Fig. 2, we show the radial distribution functions $g_{\mathrm{Me-O}}(r)$ between methane and oxygen site of water which were calculated in previous computer simulations of aqueous solution of methane.[46,52] According to Fig. 2, the peak and dip of $g_{\mathrm{Me-O}}(r)$ are shifted toward smaller distances when the temperature is raised at constant density. These shifts are then consistent with the decrease of $\rho^e(\epsilon)$ in the ϵ region more favorable than ~ -0.1 kcal/mol and the increase at the larger ϵ region. When the density is varied at a constant supercritical temperature of 400 °C, $\rho^e(\epsilon)/\rho^n$ decreases with the density reduction outside the weak interaction region of ϵ between ~ -0.1 and 0 kcal/mol. In this case, the complementary observation for $g_{\mathrm{Me-O}}(r)$ in Fig. 2 is that the density reduction leads to the decrease of $g_{\mathrm{Me-O}}(r)$ in the short-range region of $r \lesssim 4.5$ Å.

When the solute is water, Fig. 1(b) shows that $\rho^e(\epsilon)$ at the ambient state of 1.0 g/cm^3 and 25 °C involves a minimum at $\epsilon \simeq -3$ kcal/mol. This minimum is often used as a criterion for the hydrogen bond.[53] At the supercritical temperature of 400 °C, however, $\rho^e(\epsilon)$ increases monotonically with ϵ when ϵ is negative. In other words, it is evident in Fig.

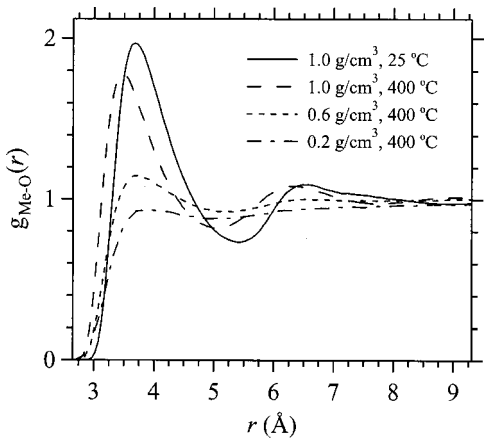

FIG. 2. The radial distribution functions $g_{\mathrm{Me-O}}(r)$ between methane and oxygen site of water as functions of the methane–oxygen distance r at an ambient state of 1.0 g/cm^3 and 25 °C and at supercritical states of 1.0, 0.6, and 0.2 g/cm^3 and 400 °C.

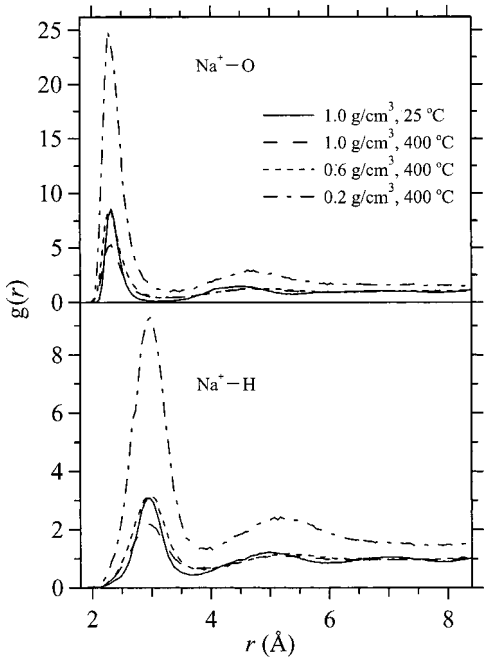

FIG. 3. The radial distribution functions $g(r)$ between sodium ion and water as functions of the distance r from the sodium ion at an ambient state of 1.0 g/cm^3 and 25 °C and at supercritical states of 1.0, 0.6, and 0.2 g/cm^3 and 400 °C.

1(b) that the solvation expressed over the energy coordinate becomes less distinct upon the temperature elevation. When the temperature is fixed at 400 °C, $\rho^e(\epsilon)/\rho^n$ in the negative ϵ region increases with the density reduction.[54] To be more precise, the increase is relatively weak for the density change from 1.0 to 0.6 g/cm^3 and is strong for the change from 0.6 to 0.2 g/cm^3. This density dependence is consistent with the complementary observation for the radial distribution function described in Ref. 45. It was seen that the radial distribution functions of pure supercritical water increase with the density reduction below $\sim$0.6 g/cm^3 and are weakly dependent on the density above $\sim$0.6 g/cm^3.[52]

$\rho^e(\epsilon)$ for sodium ion is shown in Fig. 1(c). It is observed at each thermodynamic state that a minimum is present at $\epsilon \simeq -13$ kcal/mol. Except for the persistence of the minimum, the density and temperature dependence of $\rho^e(\epsilon)/\rho^n$ for sodium ion is similar to that for water described above. The temperature elevation leads to a less distinct solvation over the energy coordinate, and the isothermal density reduction from 0.6 to 0.2 g/cm^3 at 400 °C causes a strong increase of $\rho^e(\epsilon)/\rho^n$ in the negative ϵ region compared to that from 1.0 to 0.6 g/cm^3.[54] A consistent view is of course obtained from the complementary examination of the radial distribution function. In Fig. 3, we show the radial distribution functions between sodium ion and water. It is indeed seen from the radial distribution functions that the hydration structure becomes less distinct with the temperature elevation. Furthermore, the radial distribution functions in the distance range shown in Fig. 3 increase with the density reduction at the fixed supercritical temperature, and their variation is stronger in the lower-density region.[52]

In addition to $\rho^e(\epsilon)$ and $\rho_0^e(\epsilon)$, $\chi_0^e(\epsilon,\eta)$ is needed as an input for the approximate evaluation of the excess chemical potential in this work. Since $\chi_0^e(\epsilon,\eta)$ involves two arguments, it is convenient for illustration to introduce an averaged form of $\chi_0^e(\epsilon,\eta)$. In the present paper, we set three regions E_i ($i=1,2$, and 3) of the energy coordinate ϵ and define $\overline{H}_0^e(E_i,E_j)$ as

$$\overline{H}_0^e(E_i,E_j)=\frac{\int_{E_i}d\epsilon\int_{E_j}d\eta(\chi_0^e(\epsilon,\eta)-\rho_0^e(\epsilon)\delta(\epsilon-\eta))}{\int_{E_i}d\epsilon\rho_0^e(\epsilon)\int_{E_j}d\eta\rho_0^e(\eta)}$$
$$=\frac{\int d\mathbf{x}\int d\mathbf{y}\delta(v^f(\mathbf{x})-\epsilon)\delta(v^f(\mathbf{y})-\eta)(g_{vv,0}^f(\mathbf{x},\mathbf{y})-1)}{\int d\mathbf{x}\,\delta(v^f(\mathbf{x})-\epsilon)\int d\mathbf{y}\,\delta(v^f(\mathbf{y})-\eta)}. \quad (31)$$

According to Eq. (31), $\overline{H}_0^e(E_i,E_j)$ represents the average of the correlation function $g_{vv,0}^f$ of two solvent molecules contained in the energy regions E_i and E_j. It should be noted that $\chi_0^e(\epsilon,\eta)$ and $\overline{H}_0^e(E_i,E_j)$ reflect only the two-body correlation of pure solvent for any solute–solvent interaction $v^f(\mathbf{x})$. These energy correlation functions correspond to the "projections" of a single correlation function $g_{vv,0}^f$ induced by $v^f(\mathbf{x})$ (and E_i).

As done in Fig. 1 for $\rho^e(\epsilon)$ and $\rho_0^e(\epsilon)$, we employ methane, water, and sodium ion to illustrate the behavior of $\overline{H}_0^e$. In the present paper, we take E_i ($i=1,2$, and 3) as

$$\text{methane}\begin{cases}E_1:\ \epsilon\leqslant -0.1\\ E_2:\ -0.1<\epsilon<10\\ E_3:\ \epsilon\geqslant 10\end{cases}$$

$$\text{water}\begin{cases}E_1:\ \epsilon\leqslant -3\\ E_2:\ -3<\epsilon<10\\ E_3:\ \epsilon\geqslant 10\end{cases} \quad (32)$$

$$\text{sodium ion}\begin{cases}E_1:\ \epsilon\leqslant -13\\ E_2:\ -13<\epsilon<10\\ E_3:\ \epsilon\geqslant 10\end{cases}$$

TABLE II. Averaged form $\overline{H_0^e}$ of the two-body correlation function in the pure solvent system.[a]

Solute	Variable	Thermodynamic state: 1.0 g/cm^3 and 25 °C	1.0 g/cm^3 and 400 °C	0.6 g/cm^3 and 400 °C	0.2 g/cm^3 and 400 °C
methane	$\overline{H_0^e}(E_1,E_1)$	$(-5.29\pm0.02)\times10^{-2}$	$(-5.37\pm0.01)\times10^{-2}$	$(-4.85\pm0.14)\times10^{-2}$	$(1.42\pm0.22)\times10^{-1}$
	$\overline{H_0^e}(E_2,E_2)$	$(-9.03\pm0.10)\times10^{-3}$	$(-8.82\pm0.02)\times10^{-3}$	$(-7.46\pm0.79)\times10^{-3}$	$(7.27\pm1.34)\times10^{-2}$
	$\overline{H_0^e}(E_3,E_3)$	$(-2.46\pm0.02)\times10^{-1}$	$(-2.41\pm0.01)\times10^{-1}$	$(-2.46\pm0.02)\times10^{-1}$	$(1.16\pm0.12)\times10^{-1}$
	$\overline{H_0^e}(E_1,E_2)$	$(-4.82\pm0.18)\times10^{-3}$	$(-5.27\pm0.07)\times10^{-3}$	$(-2.27\pm1.09)\times10^{-3}$	$(1.05\pm0.17)\times10^{-1}$
	$\overline{H_0^e}(E_1,E_3)$	$(-1.50\pm0.34)\times10^{-3}$	$(2.42\pm0.33)\times10^{-3}$	$(1.30\pm0.18)\times10^{-2}$	$(3.09\pm0.23)\times10^{-1}$
	$\overline{H_0^e}(E_2,E_3)$	$(-2.26\pm0.18)\times10^{-3}$	$(-2.97\pm0.09)\times10^{-3}$	$(0.30\pm1.16)\times10^{-3}$	$(1.09\pm0.18)\times10^{-1}$
water	$\overline{H_0^e}(E_1,E_1)$	$(-2.01\pm0.06)\times10^{-1}$	$(-1.90\pm0.05)\times10^{-1}$	$(-2.20\pm0.07)\times10^{-1}$	$(-0.18\pm0.66)\times10^{-1}$
	$\overline{H_0^e}(E_2,E_2)$	$(-8.74\pm0.05)\times10^{-3}$	$(-8.57\pm0.03)\times10^{-3}$	$(-7.71\pm0.67)\times10^{-3}$	$(9.19\pm1.35)\times10^{-2}$
	$\overline{H_0^e}(E_3,E_3)$	$(-3.28\pm0.04)\times10^{-1}$	$(-3.08\pm0.01)\times10^{-1}$	$(-3.19\pm0.03)\times10^{-1}$	$(3.85\pm1.84)\times10^{-2}$
	$\overline{H_0^e}(E_1,E_2)$	$(-7.66\pm0.06)\times10^{-3}$	$(-6.82\pm0.14)\times10^{-3}$	$(-4.36\pm1.13)\times10^{-3}$	$(1.33\pm0.16)\times10^{-1}$
	$\overline{H_0^e}(E_1,E_3)$	$(-6.45\pm0.17)\times10^{-2}$	$(-7.51\pm0.12)\times10^{-2}$	$(-4.57\pm0.33)\times10^{-2}$	$(3.96\pm0.24)\times10^{-1}$
	$\overline{H_0^e}(E_2,E_3)$	$(-2.31\pm0.10)\times10^{-3}$	$(-2.43\pm0.07)\times10^{-3}$	$(0.49\pm1.06)\times10^{-3}$	$(1.46\pm0.18)\times10^{-1}$
sodium ion	$\overline{H_0^e}(E_1,E_1)$	$(-4.51\pm0.05)\times10^{-1}$	$(-3.92\pm0.03)\times10^{-1}$	$(-4.76\pm0.04)\times10^{-1}$	$(-4.50\pm0.23)\times10^{-1}$
	$\overline{H_0^e}(E_2,E_2)$	$(-7.45\pm0.07)\times10^{-3}$	$(-7.35\pm0.03)\times10^{-3}$	$(-6.11\pm1.10)\times10^{-3}$	$(6.69\pm1.05)\times10^{-2}$
	$\overline{H_0^e}(E_3,E_3)$	$(-2.20\pm0.02)\times10^{-1}$	$(-2.30\pm0.01)\times10^{-1}$	$(-2.27\pm0.02)\times10^{-1}$	$(8.61\pm1.00)\times10^{-2}$
	$\overline{H_0^e}(E_1,E_2)$	$(-4.91\pm0.20)\times10^{-3}$	$(-5.10\pm0.12)\times10^{-3}$	$(-2.84\pm1.84)\times10^{-3}$	$(1.02\pm0.12)\times10^{-1}$
	$\overline{H_0^e}(E_1,E_3)$	$(1.52\pm2.12)\times10^{-3}$	$(-1.46\pm0.10)\times10^{-2}$	$(5.44\pm0.39)\times10^{-2}$	$(6.37\pm0.22)\times10^{-1}$
	$\overline{H_0^e}(E_2,E_3)$	$(-3.34\pm0.23)\times10^{-3}$	$(-2.89\pm0.08)\times10^{-3}$	$(-0.30\pm1.60)\times10^{-3}$	$(1.08\pm0.13)\times10^{-1}$

[a]The error is estimated at 95% confidence level from the six sets of computer simulations of the pure solvent system through an expression similar to Eq. (26).

where the energy is expressed in the unit of kcal/mol. This choice of E_i is motivated by the above discussion concerning Fig. 1. E_1 may be viewed as the favorable region of solvation, and E_3 as the unfavorable region corresponding to the excluded volume by the solute–solvent repulsion. E_2 is the intermediate energy region including the solute–solvent interactions at large distances. $\overline{H_0^e}$ determined under Eq. (32) is then shown in Table II. At the high-density states of 1.0 g/cm^3 and 25 and 400 °C, $\overline{H_0^e}$ is large in magnitude for the autocorrelation in the E_1 and E_3 regions. The largest cross correlation $\overline{H_0^e}(E_1,E_3)$ among the three model solutes in Table II is found for water, which involves the most delicate competition between the favorable and unfavorable contributions of the solute–solvent interaction to the excess chemical potential. On the other hand, the density reduction from 0.6 to 0.2 g/cm^3 at a fixed temperature of 400 °C leads to a strong increase of $\overline{H_0^e}(E_i,E_j)$ for each solute compared to that from 1.0 to 0.6 g/cm^3. This density dependence corresponds, of course, to the complementary observation for the radial distribution function of pure supercritical water.[45,52]

ACKNOWLEDGMENTS

N. M. acknowledges the Research Grant-in-Aid from the Ministry of Education, Science, and Culture (Grant Nos. 11740322 and 13640509). He is also grateful to Professor M. Ikeguchi of Yokohama City University for valuable discussions and suggestions and to the Supercomputer Laboratory of Institute for Chemical Research, Kyoto University, for generous allocation of computation time.

APPENDIX: NUMERICAL SCHEME OF DISCRETIZATION

When the solute–solvent interaction of interest is continuous, it is necessary to discretize the energy coordinate ϵ in the numerical realization of the energy distribution functions. In this Appendix, we describe the scheme of discretization in detail.

The first step in the discretization is to fix the minimum and maximum values of the energy coordinate ϵ. The minimum value $\epsilon_{\min}$ was simply taken to be smaller than the lowest possible value of the solute–solvent interaction energy. Of course, the choice of $\epsilon_{\min}$ has no effect as far as $\epsilon_{\min}$ is smaller than the lowest possible energy between the solute and solvent. The maximum value $\epsilon_{\max}$ of the coordinate ϵ was fixed at $10^{13}k_BT$ throughout the present work. We see below that our choice of $\epsilon_{\max}$ is large enough and does not affect the approximate evaluation of the excess chemical potential. In addition to $\epsilon_{\min}$ and $\epsilon_{\max}$, we set ϵ_{core} at which the discretization scheme is changed. When $\epsilon_{\min}\leqslant\epsilon\leqslant\epsilon_{\text{core}}$, the discretization is performed on the linear basis. In other words, an interval Δ is chosen and the energy coordinate ϵ is discretized as

$$\epsilon_i=\epsilon_{\min}+(i-1)\Delta, \tag{A1}$$

where $i=1,\cdots,L$ and

$$\epsilon_1=\epsilon_{\min}$$
$$\epsilon_L=\epsilon_{\text{core}}. \tag{A2}$$

When $\epsilon_{\mathrm{core}} \leqslant \epsilon \leqslant \epsilon_{\mathrm{max}}$, on the other hand, the discretization is carried out on the logarithmic basis. In this case, the number M of steps is chosen so that the discretized coordinate is given by

$$\epsilon_i = \epsilon_{\mathrm{core}} \left(\frac{\epsilon_{\mathrm{max}}}{\epsilon_{\mathrm{core}}} \right)^{(i-L)/M}, \tag{A3}$$

where $i = L, \cdots, (L+M)$ and

$$\epsilon_{L+M} = \epsilon_{\mathrm{max}}. \tag{A4}$$

The instantaneous distribution $\hat{\rho}^e$ defined by Eq. (1) [actually, its reduced form defined by Eq. (22)] is discretized correspondingly and is denoted by $\hat{\rho}_i^e$ $(i = 1, \cdots, (L+M))$. $\hat{\rho}_i^e$ is the instantaneous solvent density in the finite region of the solute–solvent interaction energy which is closest to ϵ_i among the $(L+M)$ discretized energies.

When the solute is nonpolar, Δ in Eq. (A1) was set to $0.001 k_B T$ and M in Eq. (A3) was 300. ϵ_{core} was chosen under the condition that it is positive and close to $0.1 k_B T$. It was actually found in this condition that the approximate calculations of the excess chemical potentials are not affected by the choice of ϵ_{core}. On the other hand, Δ was $0.05 k_B T$ for water, methanol, and ethanol and $0.1 k_B T$ for Na^+, Cl^-, and NaCl. For these solutes, M in Eq. (A3) was 100. The condition to choose ϵ_{core} was that it is a value above which the solute–solvent distribution $\rho^e(\epsilon)$ given by Eq. (5) is calculated to be zero numerically in the solution system. In this condition, the choice of ϵ_{core} was found to be indifferent to the approximate calculations of the excess chemical potentials.

The energy intervals Δ in Eq. (A1) for the nonpolar solutes were chosen to be small compared to those for the polar and ionic solutes. This is related to the long-range behavior of the solute–solvent interaction. When the solute is polar or ionic, its interaction with the solvent at large distances may be both positive and negative. Thus, although the solvent molecule interacting weakly with the solute molecule ($\epsilon \approx 0$) is large in number, its contribution to the ϵ integral of Eq. (12) is almost canceled and a fine interval is not necessary around $\epsilon = 0$. When the solute is nonpolar, in contrast, the solute–solvent interaction is always nonpositive at large distances. In this case, there is no cancellation for the contribution from the solvent molecules with small $|\epsilon|$ and the interval needs to be fine around $\epsilon = 0$.

The inputs necessary for evaluating an excess chemical potential through Eqs. (12), (18), (19), (20), and (21) are the energy distribution functions $\rho^e(\epsilon)$, $\rho_0^e(\epsilon)$, and $\chi_0^e(\epsilon, \eta)$ given by Eq. (5). The discretization of these distribution functions simply follows that of the instantaneous distribution $\hat{\rho}^e$ described above and the resulting excess chemical potentials are listed in Table I. In order to see the effect of discretization, we also define a coarse-grained version $\epsilon_{i;K}$ $(i = 1, \cdots, (L+M)/K)$ of the energy coordinate and the corresponding version $\hat{\rho}_{i;K}^e$ of the instantaneous distribution as

$$\epsilon_{i;K} = \frac{\epsilon_{(i-1)K+1} + \epsilon_{iK}}{2},$$
$$\hat{\rho}_{i;K}^e = \frac{\sum_{j=(i-1)K+1}^{iK} \delta_j \hat{\rho}_j^e}{\sum_{j=(i-1)K+1}^{iK} \delta_j}, \tag{A5}$$

where δ_j is the volume (length) of the region of ϵ which is closest to ϵ_j among the $(L+M)$ discretized energies. The distribution functions $\rho^e(\epsilon)$, $\rho_0^e(\epsilon)$, and $\chi_0^e(\epsilon, \eta)$ are coarsely discretized accordingly. It was then found that when the energy coordinate is coarsened by a factor of $K \leqslant 5$, the excess chemical potentials from the approximate scheme are unchanged within 0.4 kcal/mol for Na^+, Cl^-, and NaCl and within 0.1 kcal/mol for the other solutes. Thus, the discretization of the coordinate in the present work does not affect the precision of the approximate excess chemical potentials shown in Table I.

To assess the effect of the choice of the maximum value ϵ_{max} of the energy coordinate, we employ a shortened version $\hat{\rho}_i^{e,S}$ $(i = 1, \cdots, (L+M-S))$ of the instantaneous distribution given by

$$\hat{\rho}_i^{e,S} = \begin{cases} \hat{\rho}_i^e & (\text{when } i < L+M-S) \\ \dfrac{\sum_{j=L+M-S}^{L+M} \delta_j \hat{\rho}_j^e}{\sum_{j=L+M-S}^{L+M} \delta_j} & (\text{when } i = L+M-S) \end{cases}. \tag{A6}$$

In this case, when $S \leqslant 200$ for the nonpolar solutes and $S \leqslant 80$ for the polar and ionic solutes, the corresponding versions of the energy distribution functions in Eq. (5) were seen to change the approximate excess chemical potentials by less than 0.1 kcal/mol. Thus, ϵ_{max} adopted in the present work is large enough in the approximate evaluation of the excess chemical potential. In addition, our choice of ϵ_{max} is validated by introducing an upper cutoff into the ϵ integral of Eq. (12) and examining the cutoff dependence of the integral.

Finally, it should be noted that (the discretized version of) the average distribution function ρ_0^e in the pure solvent is sometimes calculated in the simulation to be zero numerically at an energy coordinate ϵ_I which is close to the minimum value ϵ_{min} or the maximum value ϵ_{max}.[55] The zero ρ_0^e at that ϵ_I then causes trouble in determining w^e and w_0^e through Eq. (5). The approximate calculation of the excess chemical potential was found to be unaffected within 0.1 kcal/mol, however, by using the energy distribution functions constructed from a modified version $\hat{\rho}_i^{e,m}$ $(i = 1, \cdots, (L+M-1))$ of the instantaneous distribution defined as[56]

$$\hat{\rho}_i^{e,m} = \begin{cases} \hat{\rho}_i^e & (\text{when } i < I) \\ \hat{\rho}_{i+1}^e & (\text{when } i \geqslant I) \end{cases}. \tag{A7}$$

[1] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic, London, 1986).
[2] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[3] The particle insertion method is an efficient method for calculating the chemical potential (Refs. 1 and 2). The method is not of general use, however, and is applicable only to a solution system with a weakly interacting solute.
[4] When the solute is at finite concentration, the "pure solvent" refers to a

mixed solvent system which involves the solute molecule as one of the solvent species.
[5] N. Matubayasi, L. H. Reed, and R. M. Levy, J. Phys. Chem. **98**, 10640 (1994).
[6] N. Matubayasi, E. Gallicchio, and R. M. Levy, J. Chem. Phys. **109**, 4864 (1998).
[7] D. Chandler and H. C. Andersen, J. Chem. Phys. **57**, 1930 (1972).
[8] G. Stell, G. N. Patey, and J. S. Høye, Adv. Chem. Phys. **48**, 183 (1981).
[9] F. Hirata and P. J. Rossky, Chem. Phys. Lett. **83**, 329 (1981).
[10] P. J. Rossky, B. M. Pettitt, and G. Stell, Mol. Phys. **50**, 1263 (1983).
[11] F. Hirata and R. M. Levy, J. Phys. Chem. **91**, 4788 (1987).
[12] M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).
[13] D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).
[14] T. Morita and K. Hiroike, Prog. Theor. Phys. **23**, 1003 (1960).
[15] L. L. Lee, J. Chem. Phys. **60**, 1197 (1974).
[16] S. J. Singer and D. Chandler, Mol. Phys. **55**, 621 (1985).
[17] R. Kjellander and S. Sarman, J. Chem. Phys. **90**, 2768 (1989).
[18] O. E. Kiselyov and G. A. Martynov, J. Chem. Phys. **93**, 1942 (1990).
[19] P. Attard, J. Chem. Phys. **94**, 2370 (1991).
[20] X. S. Chen, F. Forstmann, and M. Kasch, J. Chem. Phys. **95**, 2832 (1991).
[21] L. L. Lee, J. Chem. Phys. **97**, 8606 (1992).
[22] L. Lue and D. Blankschtein, J. Chem. Phys. **100**, 3002 (1994).
[23] N. Matubayasi and M. Nakahara, J. Chem. Phys. **113**, 6070 (2000).
[24] In the previous paper (Ref. 23), we used such phrases as ''energetic representation'' and ''energetic coordinate.'' In this paper, we correct the word ''energetic'' into ''energy,'' and adopt such phrases as ''energy representation,'' ''energy coordinate,'' and ''energy distribution function.''
[25] O. Kajimoto, Chem. Rev. **99**, 355 (1999).
[26] P. E. Savage, Chem. Rev. **99**, 603 (1999).
[27] N. Matubayasi, C. Wakai, and M. Nakahara, Phys. Rev. Lett. **78**, 2573 (1997); **78**, 4309 (1997).
[28] N. Matubayasi, C. Wakai, and M. Nakahara, J. Chem. Phys. **107**, 9133 (1997).
[29] N. Matubayasi, N. Nakao, and M. Nakahara, J. Chem. Phys. **114**, 4107 (2001).
[30] When the intramolecular degrees of freedom such as the molecular vibrations and intramolecular polarization are present in the solute and/or solvent molecule, the developments in the present paper are valid simply by redefining the full coordinate $\mathbf{x}$ as a collection of the variables that are enough to specify the solute–solvent interaction potential. See Appendix A of Ref. 23.
[31] In the approximate treatment developed in the present paper, it is not necessary to assume that the pure solvent system is homogeneous and isotropic. In other words, a system with an external field can be taken as the initial state of the solute insertion and be called the ''pure solvent'' system.
[32] When Eq. (19) is adopted, it is actually unnecessary to suppose a set of inequalities that $w^e(\epsilon)<0$ when $\rho^e(\epsilon)/\rho_0^e(\epsilon)<1$ or $\epsilon>0$ and that $w^e(\epsilon)>0$ when $\rho^e(\epsilon)/\rho_0^e(\epsilon)>1$ or $\epsilon<0$. As illustrated in Fig. 1, however, this set of inequalities is mostly valid. A similar set of inequalities for $w_0^e(\epsilon)$ holds mostly, furthermore, since $w_0^e(\epsilon)$ is usually found to have the same sign as $w^e(\epsilon)$ for the model systems in the present work. It should be noted that the density and temperature dependence of $w^e(\epsilon)$ at fixed ϵ is parallel to that of $\rho^e(\epsilon)/\rho^n$.
[33] In the favorable region, $\rho_0^e(\epsilon)$ is often small. It is then difficult to obtain $\chi_0^e(\epsilon,\eta)$ with good precision when both ϵ and η are in the favorable region.
[34] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[35] In the RISM integral equation approaches, it is necessary to place a repulsive core at the hydrogen site for the sake of avoiding numerical instability (Refs. 11 and 13). In our approach, it is not required that each interaction site carry a repulsive core. In the ion–water potentials employed in this work, the Lennard-Jones interaction is present between the hydrogen site and ion, just because we adopted the original formulations of the potential model.
[36] When an intermolecular interaction potential is truncated spherically [or more generally in the form of Eq. (23)], the potential is not rigorously of site–site form. The energy representation is advantageous, however, in that its applicability is not restricted by the form of the potential function.
[37] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).
[38] D. E. Smith and A. D. J. Haymet, J. Chem. Phys. **98**, 6445 (1993).
[39] J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).
[40] W. L. Jorgensen, J. Phys. Chem. **90**, 1276 (1986).
[41] B. M. Pettitt and P. J. Rossky, J. Chem. Phys. **84**, 5836 (1986).
[42] O. Steinhauser, Mol. Phys. **45**, 335 (1982).
[43] It is of interest to note that the particle insertion method has a similar feature that the calculation of the excess chemical potential of a solute does not affect the sequence of configurations in the simulation. The particle insertion method is not useful, however, when a solvent molecule is almost always found in the core region of the interaction potential with the solute molecule inserted. In the calculation of $\rho_0^e(\epsilon)$ and $\chi_0^e(\epsilon,\eta)$, on the other hand, a solvent molecule in the core region simply contributes to the distribution at a large and positive value of the solute–solvent interaction energy.
[44] Actually, the simulation of the solution system is not necessary to perform when the solute–solvent interaction is strongly unfavorable and $\rho^e(\epsilon)\approx 0$ holds over the whole range of ϵ. In this case, $\alpha(\epsilon)\approx 0$ in Eq. (18) and $w^e(\epsilon)$ is not needed in the approximate evaluation of $\Delta\mu$.
[45] N. Matubayasi and M. Nakahara, J. Chem. Phys. **112**, 8089 (2000).
[46] N. Matubayasi and M. Nakahara, J. Phys. Chem. B **104**, 10352 (2000).
[47] In Ref. 46, the indirect part (solvent-mediated part) of the potential of mean force was calculated between two methane molecules in water. The ''excess chemical potential'' of a pair of methane molecules is then simply the sum of the indirect part of the potential of mean force and twice the excess chemical potential of a single methane molecule. Note that the direct interaction between the methane molecules is not included in the calculation of the ''excess chemical potential'' of a pair of methane molecules.
[48] M. Zacharias, T. P. Straatsma, and J. A. McCammon, J. Chem. Phys. **100**, 9025 (1994).
[49] P. B. Balbuena, K. P. Johnston, and P. J. Rossky, J. Phys. Chem. **99**, 1554 (1995).
[50] P. B. Balbuena, K. P. Johnston, and P. J. Rossky, J. Phys. Chem. **100**, 2706 (1996).
[51] When the pure solvent system is homogeneous and isotropic, $g_{uv,0}^f$ depends only on the relative configuration of the two solvent molecules concerned.
[52] Under the set of potential functions employed in the present work, only the methane–oxygen radial distribution function is sufficient to determine $\rho^e(\epsilon)$ for methane through Eq. (29). When the solute is water or sodium ion, the knowledge of the site–site radial distribution functions are not enough to determine $\rho^e(\epsilon)$ through Eq. (29) or $\chi_0^e(\epsilon,\eta)$ through Eq. (30). Actually, the determination is possible only when $g_{uv}^f(\mathbf{x})$ and $g_{uv,0}^f(\mathbf{x},\mathbf{y})$ can be constructed from the set of site–site radial distribution functions. See Appendix D of Ref. 23.
[53] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[54] When the solute is water, $\rho^e(\epsilon)$ decreases with the density reduction at constant temperature. When the solute is sodium ion, on the other hand, the isothermal density reduction leads to an increase of $\rho^e(\epsilon)$ in the strongly favorable region of $\epsilon\lesssim -22$ kcal/mol and a decrease of $\rho^e(\epsilon)$ in $\epsilon\gtrsim -22$ kcal/mol.
[55] When (the discretized version of) the average distribution function ρ^e in the solution is calculated to be zero numerically at ϵ_I, w^e cannot be determined at that ϵ_I. In this case, since the weight factor $\alpha(\epsilon)$ is zero by Eq. (21), the terms involving w^e at ϵ_I do not contribute to the integral (or its discretized sum) of Eq. (12).
[56] When (the discretized version of) the average distribution function ρ_0^e in the pure solvent is calculated to be zero numerically at more than one energy coordinates, the definition of $\hat{\rho}_i^{e,m}$ needs to be modified from Eq. (A7) accordingly by skipping those energy coordinates.